

Amendments to the Drawings:

The attached two sheets of drawings include Figs. 3 and 5 on separate sheets and replace the original sheet including both Figs. 3 and 5.

Attachment: Replacement Sheets

REMARKS/ARGUMENTS

In view of the foregoing amendments and the following remarks, the applicant respectfully submits that the pending claims are statutory under 35 U.S.C. § 101, comply with 35 U.S.C. § 112 and are not anticipated under 35 U.S.C. § 102. Accordingly, it is believed that this application is in condition for allowance. If, however, the Examiner believes that there are any unresolved issues, or believes that some or all of the claims are not in condition for allowance, the applicant respectfully requests that the Examiner contact the undersigned to schedule a telephone Examiner Interview before any further actions on the merits.

The applicant will now address each of the issues raised in the outstanding Office Action.

Objections

The drawings stand objected to because the drawing sheets are presented out of sequential order (Figure 5 appears before Figure 4.) The drawings have been amended to split the drawing sheet that included Figures 3 and 5 into two separate drawing sheets. Accordingly, this objection should be withdrawn.

The abstract of the disclosure is objected to because it identified a system and method, while claim 50 also claims an apparatus. Since the abstract has been amended so that the statutory form of the invention is not specified, this objection should be withdrawn.

Claim 32 is objected to because it includes the recitation "at least one record specifying at least one such word as a key into the hash table". The Examiner notes that the term "record" does not appear in base claim 17, nor does it appear in intervening claim 31. The Examiner interpreted record to refer to words in the source or corpus text, contending that it was not found to be identified in the disclosure as a particular data type. However, a hash table "record" is a well known term of art and is clearly not the source document or text corpus. Since "record" is a well understood term of art and since its use in the claims does not raise any antecedent basis issues, this objection should be withdrawn

Rejections under 35 U.S.C. § 112

Claims 1-7 stand rejected under 35 U.S.C. § 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention. Specifically, the Examiner contends that the elements "capitalizer," "tokenizer," "processor," and "preprocessor" are defined with functions that are contradictory or mutually exclusive. The applicant respectfully disagrees.

The Examiner argues that "[t]he disclosure identifies the capitalizer as *the* element that tokenizes... [Emphasis added.]" (Paper No. 20051125, page 4.) The Examiner also notes that Figure 4 shows that tokenizer 65 is contained in the capitalizer 64

element, and that Figure 4 illustrates a processor 64 but no preprocessor to tokenize. (Paper No. 20051125, page 5.)

The applicant first notes that a "processor" is not an expressly claimed element. Second, and more importantly, the applicant respectfully draws the Examiner's attention to the lexicon builder of Figure 2 which also includes a tokenizer 36. The lexicon builder 31 is an example of a "preprocessor". In view of this clarification, the applicant respectfully submits that claims 1-7 comply with 35 U.S.C. § 112, second paragraph and that this rejection should therefore be withdrawn.

Claim 15 stands rejected under 35 U.S.C. § 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention. Specifically, the Examiner states that claim 15 is an improper hybrid claim. Claim 49 stands rejected under 35 U.S.C. § 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention for a similar reason.

Since these claims have been canceled, this ground of rejection is rendered moot.

Rejections under 35 U.S.C. § 101

Claims 15 and 49 stand rejected under 35 U.S.C. § 101 because the claimed invention is purportedly directed to non-statutory subject matter. The applicant respectfully requests that the Examiner reconsider and

withdraw this ground of rejection in view of the following.

Since these claims have been canceled, this ground of rejection is rendered moot.

Rejections under 35 U.S.C. § 102

Claims 1-50 stand rejected under 35 U.S.C. § 102(b) as being anticipated by U.S. Publication No. 2002/0099744 ("the Coden publication"). The applicant respectfully requests that the Examiner reconsider and withdraw this ground of rejection in view of the following.

Before addressing various patentable features of the claimed invention, the Coden publication is introduced. The Coden publication concerns providing capitalization recovery for text. One embodiment operates without phrase processing (phrase processing is discussed below) (See, e.g., Figure 3.), while another embodiment operates with phrase processing (See, e.g., Figure 8.). As indicated by block 50 in Figures 3 and 8, both embodiments apply preprocessing. Preprocessing defines words as any sequence of characters between a current position and a next space. Thus, "words" include punctuation immediately following it. (See, e.g., paragraph [0073].) The output of the processing is an input to the capitalization recovery techniques. (See, e.g., block 63 of Figures 3 and 8.)

In the first embodiment, each input word is capitalized depending on whether or not certain punctuation rules apply (See, e.g., 300 of Figure 3, and Figure 6.), whether or not the word is a title (See, e.g., 100 of Figure 3, and Figure 4.), whether or not the

word is an abbreviation (See, e.g., 200 of Figure 3, and Figure 5.), and whether or not the word is an entry in a single dictionary (the applicants assume this was intended to be the capitalization dictionary) (See, e.g., 500 of Figure 3, and Figure 7.) As described in paragraphs [0055]-[0063], the capitalization dictionary includes the number of times a word appears in all lower case (l), capitalized (c), in all upper case (u) and at the start of a sentence (m). A probability that the word should be capitalized can be computed from this information (See, e.g., paragraph [0061].). If the probability exceeds a threshold (e.g., 0.5), the word is capitalized. (See, e.g., paragraph [0063].) If the word is not in the capitalization dictionary, it is assumed to be a named entity and is therefore capitalized. (See, e.g., paragraph [0064].) As can be appreciated from the foregoing, the first embodiment applies a series of heuristic (e.g., punctuation-based rules), statistical (e.g., capitalization probability), and dictionary-based (e.g., titles and abbreviations) processing to effect capitalization recovery. (See, e.g., paragraph [0094].)

In the second embodiment, words are processed using a singleton dictionary and a phrase dictionary. (See, e.g., Figures 8, 10, and 11.) The singleton and phrase dictionaries are generated using capitalization probabilities, each of which is based on the number of times a word or phrase appears in all lower case (l), capitalized (c), in all upper case (u) and at the start of a sentence (m). Those words or phrases with a capitalization probability greater than a predetermined value (which is denoted as a filtering act 1240 in Figure

12) are stored in the singleton or phrase dictionaries.
(See, e.g., paragraphs [0092] and [0093].)

As shown in Figure 10, records of the singleton dictionary each include the term 1020, the minimum length of a phrase (which may be subject to an occurrence threshold) which the term begins 1030, the maximum length of a phrase (which may be subject to an occurrence threshold) which the term begins 1040, and perhaps a preferred spelling (other than default capitalization with only first letter of the word capitalized) of the word 1050.

As shown in Figure 11, records of the phrase dictionary each include the phrase 1120 and perhaps a preferred spelling (other than default capitalization with only the first letter of the first word of the phrase capitalized) of the phrase 1130.

If a word being processed is found in the singleton dictionary, assuming that n is the maximum length of the phrase which the word begins, the next $n-1$ words of the input being processed are joined with the word and looked up in the phrase dictionary. If the words are found in the phrase dictionary, and there is a preferred spelling, the preferred spelling is applied to the phrase. If the words are found in the phrase dictionary, but there is no preferred spelling, the default spelling is applied to the phrase. If the words are not found in the phrase dictionary, the phrase under consideration is shorted by one word and the process is repeated until the phrase is found in the phrase dictionary (or until the number of words is 1). (See, e.g., paragraphs [0087]-[0089].)

Claims 1-14 and 16

Independent claims 1, 8 and 16, as amended, are not anticipated by the Coden publication because the Coden publication does not disclose an act or element which analyzes the set of words for correct capitalization, including skipping at least one word of a set if it is determined to be of a predefined type such that the capitalizer does not capitalize the at least one such word. The Examiner cites Figure 12, claims 10 and 19, and paragraphs [0017], [0067] and [0093] of the Coden publication as teaching this features. (See, e.g., Paper No. 20051125, page 8.)

The applicant respectfully notes that the filtering described in the cited portions of the Coden publication pertains to determining whether or not a word (or phrase) is to be added to a singleton dictionary (or phrase dictionary). More specifically, information in a capitalization dictionary can be used to determine a capitalization probability, and words (or phrases) are added to the singleton dictionary (or phrase dictionary) on the basis of the probability. (See, e.g., Figure 12 and the corresponding description.) Referring to Figure 1, the singleton dictionary 15A may be used by a singleton subsystem 500 and the phrases dictionary 15B may be used by a phrase subsystem 800. As can be appreciated from the foregoing, filtering acts performed when building a dictionary, which dictionary is later used to determine how to apply capitalization, are performed *before* acts pertaining to applying capitalization. That is, the Coden publication does not teach *skipping* at least one word of a set *if it is*

determined to be of a predefined type such that the capitalizer does not capitalize the at least one such word.

Thus, independent claims 1, 8 and 16 are not anticipated by the Coden publication for at least the foregoing reason. Since claims 2-7 depend from claim 1 and since claims 9-14 depend from claim 8, these claims are similarly not anticipated.

Further, dependent claim 2, as amended, also recites a document title capitalizer. The Examiner cites paragraph [0042] of the Coden publication as teaching a titles dictionary or list. (Paper No. 20051125, page 9.) In the Coden publication, the title dictionary 15D pertains to peoples' titles such as Dr., Gov., Mr., Ms., Rev., etc. On the other hand, the claimed invention pertains to applying capitalization to document titles. Thus, claim 2, as amended, is not anticipated by the Coden publication for at least this additional reason.

Furthermore, dependent claims 4 and 11 further define the "predefined types" of word(s) skipped. The Examiner argues that paragraph [0043] of the Coden publication teaches parsing words including numbers and words consisting entirely of consonants, and concludes that "[i]t is inherent in the Coden capitalization system that numbers [are] skipped and not capitalized because numbers are incapable of being capitalized." Paper No. 20051125, page 9. However, the claims recite that the word skipped may be one that *comprises (e.g., includes)* numbers, not solely consisting of numbers. Applying the Examiner's logic, it would not be inherent not to capitalize a word including numbers since the word might also include letters. Accordingly, claims 4 and 11 are

not anticipated by the Coden patent for at least this additional reason.

Claims 17-48 and 50

Independent claims 17, 33 and 50, as amended, are not anticipated by the Coden publication because the Coden publication does not disclose an act or element which selects *at least two capitalization variations* within the identified word set having a non-standard capitalization, and *adding* the *at least two such variations* to a lexicon. For example, referring to Figure 3 of the present application, a word set may include a word 52, and two or more {non-standard capitalization 54, frequency of occurrence} 55 pairs. On the other hand, the singleton and phrase dictionaries in the Coden patent are shown as including a *single* preferred spelling. (See, e.g., 1050 of Figure 10 and 1130 of Figure 11.) Thus, independent claims 17, 33 and 50, as amended, are not anticipated by the Coden publication for at least this reason. Since dependent claims 18-32 depend, either directly or indirectly, from claim 17 and since claims 34-48 depend, either directly or indirectly, from claim 33, these claims are similarly not anticipated by the Coden publication.

Further, dependent claims 22 and 38 also recite an act or element which normalizes a plurality of words extracted *relative to a source of* the unstructured content. As described in the specification, this is useful to prevent the contribution from any one source from dominating the lexicon, such as could occur if a content provider included a large corpus containing

improperly capitalized words. The Examiner argues that the Coden publication teaches filtering items with a high likelihood of being erroneous. (See, Paper No. 20051125, page 15.) Even assuming, arguendo, that this is true, it does not teach the specific claim language -- normalizing... relative to a source...." Accordingly, these claims are not anticipated by the Coden publication for at least this additional reason.

Further, dependent claims 23 and 39 have been amended to clarify that the set comprising significant statistics comprises only non-standard capitalization variations having at least four occurrences of at least one such variation within a word set. The Examiner reasons that since the Coden publication teaches counting the number of times a word appears in certain forms, it inherently will count up to four or more. (See Paper No. 20051125, page 16.) Even assuming, arguendo, that this is true, claims 23 and 39 as amended clarify that the count is used to define which non-standard capitalization variations are considered to have significant statistics. Accordingly, the claims are not anticipated by the Coden publication for at least this additional reason.

Further, dependent claims 28 and 44 also recite that implicit rules for capitalization comprise at least one of a number, having no vowels, and constituting at least one of an article, conjunction and preposition. The Examiner makes a general allegation that the Coden publication **separately teaches** words consisting entirely of consonants and rules for adding new words to dictionaries. Such purported separate teachings do not teach the elements as recited in the claims.

The Court of Appeals for the Federal Circuit ("the CAFC") has instructed that to anticipate, a single prior art reference must "describe all of the elements of the claims, *arranged as in the [claim].*" (Emphasis added.) C.R. Bard Inc. v. M3 Systems, Inc., 48 U.S.P.Q.2d 1225, 1230 (Fed. Cir. 1998), cert. denied, 119 S. Ct. 1804 (1999). This is in accord with previous Court of Claims and Patent Appeals ("the CCPA") decisions. For example, the CCPA has instructed that to anticipate:

[the] reference must clearly and unequivocally disclose the claimed [invention] or direct those skilled in the art to the [claimed invention] *without any need for picking, choosing and combining various disclosures not directly related to each other* by the teachings of the cited reference. [Emphasis added.]

In re Arkley, 172 U.S.P.Q. 524, 526 (CCPA 1972).

In this instance, even assuming, arguendo, that the Coden publication teaches the separate features alleged by the Examiner, these separate features are not described as being arranged as in the claim. Furthermore, the claims pertain to skipping at least one capitalization variation based on the explicit rules, while the portion of the Coden publication relied on by the Examiner teaches rules for adding new words to the dictionary. Thus, claims 28 and 44 are not anticipated by the Coden patent for at least this additional reason.

Further, dependent claims 29 and 45 are also not anticipated by the Coden patent for the additional reasons discussed above with reference to claims 22 and 38.

Further, dependent claims 30 and 46 also recite and act or element which accommodates multiple forms of capitalization for each variation by *annotating each capitalization form with a frequency count* (See, e.g., Figure 3.) and skipping those variation(s) that occur infrequently. The Examiner alleges that the Coden publication teaches protecting dictionaries from infrequently occurring erroneous entries. (Paper No. 20051125, page 19.) Even assuming, arguendo, that this is true, it does not teach accommodating multiple forms of capitalization for each variation by *annotating each capitalization form with a frequency count*. Thus, claims 30 and 46 are not anticipated by the Coden publication for at least this additional reason.

Amendments to the Specification and Drawings

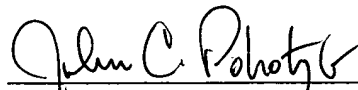
Amendments to the specification have been made, and changes to the drawings have been proposed, to correct a number of minor errors and to properly sequence the drawing figures.

Conclusion

In view of the foregoing amendments and remarks, the applicant respectfully submits that the pending claims are in condition for allowance. Accordingly, the applicant requests that the Examiner pass this application to issue.

Respectfully submitted,

July 5, 2006



John C. Pokotylo, Attorney

Reg. No. 36,242

Tel.: (732) 542-9070

CERTIFICATE OF MAILING under 37 C.F.R. 1.8(a)

I hereby certify that this correspondence is being deposited on July 5, 2006 with the United States Postal Service as first class mail, with sufficient postage, in an envelope addressed to Mail Stop Amendment, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.



John C. Pokotylo

36,242
Reg. No.